



# How to describe and evaluate “deception” phenomena: recasting the metaphysics, ethics, and politics of ICTs in terms of magic and performance and taking a relational and narrative turn

Mark Coeckelbergh<sup>1</sup> 

© The Author(s) 2017. This article is an open access publication

**Abstract** Contemporary ICTs such as speaking machines and computer games tend to create illusions. Is this ethically problematic? Is it deception? And what kind of “reality” do we presuppose when we talk about illusion in this context? Inspired by work on similarities between ICT design and the art of magic and illusion, responding to literature on deception in robot ethics and related fields, and briefly considering the issue in the context of the history of machines, this paper discusses these questions through the lens of stage magic and illusionism, with the aim of reframing the very question of deception. It investigates if we can take a more positive or at least morally neutral view of magic, illusion, and performance, while still being able to understand and criticize the relevant phenomena, and if we can describe and evaluate these phenomena without recourse to the term “deception” at all. This leads the paper into a discussion about metaphysics and into taking a relational and narrative turn. Replying to Tognazzini, the paper identifies and analyses two metaphysical positions: a narrative and performative non-dualist position is articulated in response to what is taken to be a dualist, in particular Platonic, approach to “deception” phenomena. The latter is critically discussed and replaced by a performative and relational approach which avoids a distant “view from nowhere” metaphysics and brings us back to the phenomena and experience in the performance relation. The paper also reflects on the ethical and political implications of the two positions: for the responsibility of ICT designers and users, which are seen as co-responsible magicians or co-performers, and for the responsibility of those who influence

the social structures that shape who has (more) power to deceive or to let others perform.

**Keywords** ICTs · Deception · Magic · Performance · Ethics · Politics

## Introduction

Many contemporary ICTs seem to afford experiences and raise ethical questions that are often framed with terms such as “virtual reality”, “illusion”, and “deception”. In particular, computer games, virtual reality, and augmented reality technologies are seen as creating the illusion of a different, “virtual” world, and some robots—especially so-called “social robots”—are seen as deceiving users into thinking that they are real persons, that they are companions, that they are animals, that they can speak, that they can feel, and so on.

For instance, there are now many machines that speak, such as speaking robots and speaking computer programs and apps. Consider robots such as Nao or Pepper (Softbank), apps such as Google Assistant or Siri (Apple), or new devices such as Amazon Echo’s Alexa. These machines and interfaces create the illusion that one speaks with a person. Is this ethical? Does it constitute cases of deception? Virtual reality technologies are also getting more common. In a recent interview in *New Scientist*, Metzinger warned that ‘In VR environments, we can be fooled into thinking that we are our avatars.’ (Ananthaswamy 2016). Commenting on Second Life, Pasquinelli (2010) has identified three illusions in VR: ‘the illusion that the artificial world is real, the illusion of non-mediation, and the illusion of being in the virtual environment.’ (p. 201) She has raised the question if and how this illusion is ethically different from other kinds

✉ Mark Coeckelbergh  
mark.coeckelbergh@univie.ac.at

<sup>1</sup> Department of Philosophy, University of Vienna, Vienna, Austria

of make believe such as children games (p. 205) In discussions about games, there is the question if violence in games is ethically problematic, and, more generally, whether digital games are mere entertainment or can change people's dispositions (see for instance Sparrow et al. 2015)—questions which can, and are often, related to issues regarding illusions and reality. And several authors in the field of robot ethics and machine ethics have argued that it is ethically problematic to give robots to children or to elderly people since this would create the illusion that the robot is a real nanny or parent or that it really is an animal that needs care. For instance, in the context of robot ethics in health care, Sparrow and Sparrow (2006) have argued the use of robots as companions or carers is akin to deception, since 'robots are premised on people believing that robots are something that they are not' (p. 148) and since 'failure to apprehend the world accurately is itself a (minor) moral failure' and since our well-being is not served by us merely believing that we are cared for, loved, etc., 'when in fact these beliefs are false' (p. 155). In other words, the argument is that this use of robots deceives, and deception is morally wrong since we should have an accurate view of the world and well-being is only served by true care, true love, and so on; we should not be deceived about 'the robots' real nature' (p. 155). Sharkey and Sharkey (2010) have argued that use of robots as nannies is deceiving children. While they argue that deception is not harmful in itself—they give the example of the puppeteer who 'creates the illusion that the puppets are interacting with each other and the audience' (p. 172)—it is harmful 'when it is used to lure a child into a false relationship with a robot and when it leads parents to overestimate the capabilities of a robot' (p. 173). Furthermore, while recognizing benefits of robots such as Paro for social experience (Kidd et al. 2006), Turkle has criticized 'our culture of simulation' (Turkle 2010). In *Alone Together* (2011) she has reported experiences people have with robots, and has argued that robots are designed in such a way that they trigger us to 'fool ourselves' (Turkle 2011, p. 20). Computers pretend to understand us, but what they do is mere 'performance' (p. 26). Robots only give us the illusion of meaning; 'they don't mean anything at all' (p. 124). We seek love, but our machines only give us 'performances of love' (p. 138). They offer 'the performance of emotion' (p. 286). It is also interesting in the light of the subsequent discussion in this paper that she starts her book with a quote from Plato's *Republic*: 'Everything that deceives may be said to enchant.' (Plato quoted in Turkle 2011, front matter). Clearly, deception seems a central issue here.

It is hard not to be sympathetic to, and agree with, these descriptions and evaluations. Who wants a false view of the world? Who wants fake love? At first sight, their views seem very reasonable and in accordance with common intuitions, which I share. However, in this paper I engage in the philosophical exercise of questioning the assumptions that a

discussions in terms of "deception", "the real", and similar notions is the best way to conceptualize, understand, and evaluate these problems. The paper aims not to provide a different answer, but to reframe the question. In particular, I investigate (1) if it is possible to take a more positive or at least morally neutral view on deception, illusion, and performance, *without losing the possibility of critical evaluation*, and what such a position would look like, (2) if it is possible to describe and evaluate these phenomena and experiences without reference to the language of deception at all and without involving a metaphysics of the real, and (3) what it would mean to take a relational and narrative approach to the problem.

In addition, I aim to draw some conclusions for the responsibility of designers, but also the responsibility of users. With regard to the former, let me note already at this point that the phenomena under discussion here are not always intended by designers. As has often been observed, illusions and deceptions also occur with technologies that are not designed to create such illusions and deceptions, for instance the Roomba (Sung et al. 2007; see also Scheutz 2011) or Packbot (Carpenter 2016). Yet this does not make the challenges regarding deception phenomena and the design of ICTs less urgent. On the contrary, it raises the question which I ask later on in this paper even more urgently: given these phenomena, how can designers take up responsibility for their design and its intended or non-intended illusions?

However, let me first explain my starting point in terms of approach. One way to respond to these developments and phenomena, both on the part of computer researchers and on the part of philosophers, is to take immediately a normative and defensive position, and to say that computer science and related fields *are* not about creating illusions and/or *should* not aim at this. The mentioned ethical positions go into that direction in so far as their arguments are based on the assumption that creating illusions is bad in the context of social robotics, or at least in the specific cases mentioned. While of course their work cannot and should not be reduced to this assumption and most authors recognize that there are also advantages to deception, they seem to share this point about deception and its link with thinking in terms of reality/illusion.

However, to say that ICTs and robotics—in particular social robotics—sometimes creates deception and illusion seems not to contribute much to understanding the phenomena. If what goes on here is deception, what would be a non-deceptive design and use of this technology? What exactly do we mean by "deception" in this context? Moreover, taking such a position may be perceived as accusing roboticists of unethical behaviour. This is not very helpful with regard to trying to collaboratively find out how we—people from the humanities and social sciences as well as robotics

researchers—can improve the ethical quality of design and use of ICTs.

In this paper I explore what I believe to be a more constructive and potentially more productive route: to pause the normative firework and first try to better understand the phenomena. This can be done in various ways. For instance, it is common within the human-robot interaction literature to refer to anthropomorphization and animism. Scheutz, for instance, has argued that perceived autonomy, in combination with the human tendency to anthropomorphize, results in unidirectional emotional bonds between humans and social robots, in which humans form a “relationship” with the robot but the robot lacks any capacity to do so, indeed does not care at all about humans. This, Scheutz argues, opens up many doors to exploitation (Scheutz 2011) and Musial (2016) has argued that animation is part of magical thinking which plays a role in empathetic relations with robots. However, it is questionable if labels such as “anthropomorphization”, “animism”, and even “magical thinking” used in this way really help to better understand the relevant phenomena: they seem to merely redescribe the phenomena without adding much new insights (Musial may be an exception, in so far as there is an attempt to link to a broader discussion about culture and religion). Furthermore, description of phenomena by means of psychological terms and theory, as a modern-scientific gesture, takes a lot of distance from the concrete relation between robot and user/observer, which according to the relational approach I will introduce later in this paper is problematic. Finally, clearly this argument relies once again on deception: robots only pretend to care, we use them as a source of meaning but they are not, etc. But what is this deception and illusion, and why exactly is it bad? It is worth further examining the meaning of deception and illusion, which can be done by philosophical conceptual work. In this paper I proceed by further exploring similarities between the design of these ICTs and (other) ways of creating illusions, and by using this as a basis for philosophical and ethical reflection.

In doing so, I shall not discuss general links between magic and technology, but focus on design and use of ICTs. For instance, in anthropology Gell (1994) has argued that magic is a “commentary” on ‘technical strategies in production, reproduction, and psychological manipulation’ (p. 8) and that magic is an ‘ideal technology’ in the sense that it does not have the costs and disadvantageous of real production such as struggle and effort (p. 9). Elsewhere he also pointed out that technology has the power of enchantment (Gell 1994). This work is relevant to understanding and evaluating contemporary ICTs, which aim at reducing those costs and which is often used to enchant the world (see also Coeckelbergh 2017). I will return briefly to Gell’s work when discussing the political implications of the approach I will develop. But within the limited space of this paper I

shall focus on responding to work that directly addresses design of ICTs and its link to magic and illusionism, in particular Tognazzini’s seminal paper on the topic.

First, I show how this direction of research can be supported by existing work within the field of ICT, including human–computer interaction, which already helpfully explored similarities between what illusionists and magicians do, on the one hand, and what developers of ICT systems, human–computer interfaces, etc. do. This sometimes includes asking ethical questions. I also point to a long history of machines and computing, which always included stage performances and acts showing off the “magic” of new technologies, ranging from the “Mechanical Turk” in the late eighteenth century to Apple presentations and advertisement video’s on social media today. Moreover, there are many contemporary technologies such as virtual reality technologies that directly and openly aim at creating an illusion.

Second, I distinguish between on the one hand a Platonic position which maintains a strict distinction between reality and illusion, and on the other hand positions that are critical of this metaphysics and take a different approach. I show that these different positions lead to different ethical questions. I pay particular attention to, and articulate and construct, an approach that takes seriously the pervasiveness of technology, rejects Platonic dualism regarding the question about deception (indeed the Platonic *framing* of the question in terms of deception), uses concepts of process, performance, and narrative to understand the relations between designers, users, and technology (and hence, in contrast to Turkle, uses the term performance in a morally neutral sense), considers a non-modern position according to which animism is not necessarily problematic, and asks about the role of user, which often remains relatively passive in standard models of magic as stage performance. I argue that, if combined with a relational view, this direction leads to an approach centred on understanding the experience *in* the performance (possibly co-performance) and its consequences, which avoids a distant “view from nowhere” and discussions about the properties of the robot, the metaphysics of the real, and so on.

Third, I reflect on the implications of this analysis for the responsibility of ICT designers and users. What does it mean for their ethical responsibility to see for instance roboticists as illusionists? And what does it mean for the ethical responsibility of the *user*? I argue that designers have a responsibility for designing the role and narrative related to their artefact, and indeed for designing the (co-) performance, but that the performance and the narrative created are also the responsibility of the user(s) since they also co-create it. This means that the design should be “open” enough to enable a role of the user as co-responsible co-magician. Moreover, keeping in mind phenomena such as attachment to Roomba, I argue that both users and designers

should take into account that there may be unintended consequences of their authorings and performances, which need to be revealed and evaluated. I reflect on the role of technology in these performances, emphasizing that it involves humans and non-humans.

Finally, I explore the implications of this approach for politics: what would a politics of deception and a politics of performance mean? I argue that some have more power than others to deceive and to perform (or let others perform for them), and that these questions should be discussed in relation to the democratic ideal. I also ask if perhaps the term deception can be saved, especially in discussions about the design of machines.

### Computer science, robotics, and the art of magic

While some people in the fields of computer science, robotics, and AI may be resistant to thinking of their research in terms of the art of magic and illusionism—indeed in terms of arts and crafts at all—this need not be the case at all. In fact, many people in the field have actively explored this link and have drawn conclusions for design. An interesting and seminal paper in this area is Bruce Tognazzini's 'Principles, Techniques, and Ethics of Stage Magic and their Application to Human Interface Design' (1993), which makes a direct link between what happens in stage magic and in the design of human–computer interfaces.

Tognazzini, an experienced software designer and consultant who worked for Apple, Sun Microsystems, and WebMD, claims that there is a long tradition of designing and presenting illusions, that it is a craft and act that works with apparatus, and that the design of human–computer interfaces can learn from its principles. Particularly helpful for the present discussion is his analysis of magic as involving two acts and realities:

Actually, there are two simultaneous acts performed in magic: the one the magician actually does—the magician's reality—and the one the spectators perceive—the spectators' reality: The magician's reality consists of all the sleights of hand and manipulation of gimmicked devices that make up the prosaic reality of magic. The spectators' reality, given a sufficiently competent magician, is entirely different: an alternate reality in which the normal laws of nature are repeatedly defied, a reality where the magician, as well as his or her tricks, appear supernatural (Tognazzini 1993, p. 357)

To create this alternate reality, the magicians needs knowledge of psychology and needs to have technique and technologies. Techniques include simulation and

dissimulation, careful handling of objects and attention to detail, and the manipulation of time and space. Tognazzini gives examples of how Macintosh programmers used these techniques, or at least can be interpreted as having used these techniques. We can easily expand his examples to today's speaking devices, human-like robots, and communicative apps: they too create an illusion, for instance the illusion of personhood, but in reality it's a "trick", and knowledgeable and skilled developers of technology know the tricks, the techniques, the way things really work—literally "behind the screen". They know the craft of code writing and the know-how of hardware development. As magicians and illusionists, they create interfaces and hide the real, technological face of the machine. What is required ethically, according to Tognazzini, is that they are honest about this. (p. 361) (I will return to this.)

This use and creation of technologies as a way to create illusions has a long history, which is as much a history of inventions as it is a history of shows and performances. The illusion is created during use, and is especially strong during the first use or presentation of the device. Consider for instance some examples from the history of machines, robots and automata. In ancient times, machines were already used to bring actors playing god onto the stage or technology was used to create moving and talking statues. There were already mechanical automata in ancient Greece and in medieval times. Japan had automata in the seventeenth century. In eighteenth century Europe, automata were used in performances to create illusions, for instance flute playing, eating, and chess playing (the so-called "Turk"). And today new robots or smart devices are often presented in a way that is very similar to such performances and magic shows. But not only robots and automata were and are "magic"; all kinds of ICTs are used and presented in this way, and especially when they are new they seem magic. To extend Tognazzini's Macintosh example from use to the *presentation* of devices: from the 1970s until now, Apple has always done shows, performances, to show off new technology and the "magic" of Apple computers. Consider for instance the 1984 presentation of the Macintosh which featured a speaking computer and a graphical interface, or the presentation of the iPhone in 2007, which was also a *performance* by magician-innovator Steve Jobs. Today many advertisements for new high tech products play on the magic of their devices. Moreover, returning to *use*, it is clear that today many new devices have the explicit goal of creating illusion, such as virtual reality (VR) and augmented or alternate reality technology. (Below I will give the example of Pokémon Go.)

More researchers in computer science have studied deception and magic, and have also pointed to productive uses of the lessons of magic and illusionism for the design of ICTs. For instance, Marshall et al. (2010) argued that whereas some kinds of deceptions may be unethical, in interactive

performances of computing there is ‘a need to create a sense of magical illusion as part of an entertaining and engaging user experience’ (p. 567). They focused in particular on ‘magical interactions’, on people that trick one another by using computers, involving misdirecting attention and creating false expectations. They proposed to broaden the human–computer interaction agenda ‘to consider the currently unfamiliar idea that the active deception of one user by another can be a valid approach to interaction design’ (p. 576). Furthermore, like Rowe (2007) and Adar et al. (2013) they also point to what may be considered to be an ethical use of deception: computer systems that create deceptions in order to maintain security. Hence Adar et al. suggest that there may be ‘benevolent deception, deception aimed at benefitting the user as well as the developer’ (Adar et al. 2013, abstract).

I agree with the authors that more needs to be said about (the ethics of) deception, and that deception should not be a taboo in the field. But from a philosophical point of view, the distinction between two realities which seems to underlie the deception discussion needs further elaboration and criticism.

## Reality, illusion, and more: two metaethical and ethical positions

### Two realities and the ethics of honesty

We could interpret Tognazzini’s distinction between two realities, a real one (the reality of the magician) and a fake, illusory one (the reality of the spectator) as a familiar metaethical position that goes back at least as far as Plato: there is reality versus illusion, there are appearances versus the real. Plato’s metaphor and myth/narrative of the cave, which he famously presented in his *Republic* (514–520a), is applicable here, with the magician in the role of the artist-craftsman creating the illusions by using all kinds of objects (wall, fire, objects) and in the role of the all-knowing philosopher who can distinguish between appearances and reality. The prisoners, by contrast, are the spectators who live in illusion. The only difference with Plato’s metaphor seems to be that at least some spectators of stage magic have the desire to find out about reality, want to understand the tricks. (Note that this metaphor of the prisoner also explains why there is often an adversarial relationship between magician and spectator, as recognized in the literature mentioned above: the spectator wants to find out, but the magician forces her magic onto the spectators. This could be seen as an issue concerning power, which I will explore at the end of my paper).

The normative implication of such a position, then, is the imperative to keep reality and illusion separate. Perhaps it is fine to be a “prisoner” of the magician during the show, but one should know that it is only an illusion. And indeed

Tognazzini proposes an ethics of honesty: during the show spectators are provided with an illusion and should think that the magician is supernatural, but outside the theatre magicians do not claim to be supernatural. In other words, we should maintain a distinction between illusion and reality:

the magician is not supernatural; the character he plays is. The computer is not capable of human intelligence and warmth; the character we create is. People will not end up feeling deceived and used when they discover, as they must ultimately, that the computer is nothing but a very fast idiot (Tognazzini 1993, p. 361)

For use of ICTs, this position thus means that users of for instance robots, talking machines, virtual reality technology, and games, should be made aware—through design, advertisement, and other means—that what they experience is an illusion, created by science and technology. It is fine to create illusion, for instance through a virtual reality device, as long as it is clear to the user *that* it is an illusion. Devices need then be designed and promoted in such a way that the user gains or retains this awareness, even if during use of the device this awareness may be temporarily suspended. Compare again to the show of the magician: outside the theatre, we know that it is a show and that it is deception. Thus, we can conclude from this position that the magician or designer should work in such a way that the user is provided with an illusion, but at the same time knows that the reality created by the computer is not real or that the computer, robot, app, etc. is technology, not a person.

This ethical position is in line with positions that see technology is a mere instrument and *should* be a mere instrument. For instance, Bryson (2010) argued robots are not persons, they are there to work for us, they are slaves. Consider also again the view I mentioned before by referring to the Sparrows: that robots should be designed as what they are, rather than pretending to be what they are not. This seems a reasonable position. Tognazzini’s interpretation of what happens in stage magic and illusionism is compelling, as is his application to interface design and—by expansion—to the design of ICTs. His ethics is also rather attractive since it is in line with many ordinary intuitions we have about technologies such as robots.

However, I am afraid we must complicate matters now. The rationale for doing so is twofold. First, when we look at experience and use of ICTs, we see that the phenomenology of this use and experience is sometimes difficult to describe in dualist terms, in terms of two realities. When we use the internet (e.g. through our smartphones), play games, talk to robots, etc. it seems rather than what we call “real” and “virtual” are mixed. Floridi (2015), for instance, has used the term “onlife” for emphasizing how it becomes increasingly difficult to describe our use of internet as being about “online” versus “offline”. Similarly, we could say that when

we interact with smart devices, human-like autonomous robots, voice interfaces such as Google Assistant, and so on, we do not generally experience this as “illusion” in contrast to “reality”. Now according to the Platonic position, this only shows our imprisonment in appearances. It shows how the companies developing these tools manage to give us the fake as opposed to the real. It is a lie. While this argument seems compelling, and can again be combined with an ethics of honesty (which could be developed in terms of a virtue ethics), it is not entirely satisfactory. Is our experience that there is only *one* experience and reality when we use these devices, that they are part of and entangled with our lives, entirely misguided? Or shall we at least also consider other ways to conceptualize what is going on, while still being able to criticize certain phenomena such as attachment to machines?

Second, in the history of philosophy, especially the history of twentieth century philosophy, we see that the Platonic position is far from uncontested. There are all kinds of other, less dualist or nondualist metaphysics available. Perhaps there are only appearances, perhaps there is no-one behind the mask and no “real” behind the curtain. Perhaps there are multiple realities, or different perspectives, or different levels of description, or different levels of abstraction (in the context of thinking about ICTs see Floridi’s philosophy of information, for instance). Or perhaps there is just one reality—natural or informational, for instance. These are all distinct metaphysical positions which are highly relevant to discussions about the ethics of ICTs. We can connect them to discussions in the history of philosophy, for instance anti-Platonic positions in Nietzsche, Dewey, and so on. In this paper, I do not have the space to discuss these at length, let alone to engage with the history of contemporary philosophy to which these positions must be related. For my purposes, it will suffice that I try to construct a plausible non-dualist alternative to the Platonic position articulated above, a working approach so to speak. This approach is influenced by, and will be connected and integrated with, relational ways of thinking in contemporary machine ethics (Coeckelbergh 2012, 2014; Coeckelbergh and Gunkel 2014), and further applies process and performance oriented thinking (Coeckelbergh 2017) to thinking about technology. Then I will discuss its ethical implications.

### **Magic times, performances with machines, and narrative game technologies: an alternative, non-platonic position using the concepts process, narrative, and performance**

Let me construct an alternative position by using three concepts: process, narrative, and performance.<sup>1</sup>

<sup>1</sup> The position is influenced by thinkers in 20th century philosophy who have questioned dualism such as Dewey and Wittgenstein; it is also influenced by Bergson and by Ricoeur. Latour and Pickering are

First, the Platonic metaphysics, at least as presented above, is remarkably static. If we describe what goes on in stage magic and in use of ICTs in terms of two realities, what is left out is *process*. As becomes clear from descriptions of stage magic in the literature cited above, illusionism is a temporal affair, it takes place in time, and it is even a particular configuration of time, in the sense that there is the experience time of the spectator (in a dualist framework called “subjective” time or *durée* in Bergson’s terms) and the experience time of the magician/designer/programmer/hacker etc. (“objective”, scientific time). If we shed dualistic thinking, however, we simply have different times and experiences that intersect (or not), without giving necessarily ontological priority to one of them. For Tognazzini, magicians manipulate time in the following way:

Magicians use two techniques to offset the actual time a trick (the essential working of the apparatus) takes place from the time of the spectators think it takes place: Anticipation, where the magician does the trick early, before spectators begin looking for it, and Premature Consumption, where the magician does the trick late, after spectators assume it has already occurred (Tognazzini 1993, p. 359)

Thus, there are indeed different times: the time of the magician and the time of the spectator. With regard to ICTs, this means that there is the time and timing of the program known by the designer and there is the time and timing of the user. But instead of seeing these different times in terms of what “really” goes on versus what is illusion, or instead of ‘offsetting time of reality from time of illusion’ and speaking in terms of ‘actual’ time versus apparent time, as Tognazzini does (p. 359), we could see these different times as belonging to *one* reality, not understood as a static world but as a process or a combination of processes.

Second, in order to move beyond dualistic thinking about these experiences but still distinguish between the experience and actions of designers and those of users/spectators, one could also talk about two *narratives*, which may or may not interlock at different times. There is the narrative of the magician/designer, including a plot with a character (the magician as artist, craftsman, scientist, and so on) and events happening (e.g. the coin is moved into the pocket of one’s jacket). There is also the narrative of the spectators,

Footnote 1 (continued)

also to be mentioned. However, for the purpose of this paper I will focus on the articulation and construction of my working alternative position in response to Tognazzini rather than on interpretation of their work, and only mention some of these authors in passing when directly relevant for the argument.

which in the cases of “deception” under consideration in this paper tends to differ from the narrative of the magician/designer, but also has a narrative structure which involves a plot with characters, including the magician *as* magician, as a supernatural being [indeed Tognazzini says that the magician plays a ‘character’ (p. 361)], and events such as the disappearance of a person. In a dualist framework, this play is put in terms of the “real” narrative versus the “illusory” narrative. But one could also see two narratives, without giving one ontological priority.

Moreover, the two narratives may be entangled to some extent and in any case they are connected. If a person uses a robot as companion, the narrative of personal companionship and the narrative of the computer program running the robot may be very different, but in practice, in use and experience, they are connected. Sometimes narratives merge, as in augmented reality or alternate reality games. Consider for instance the game Pokémon Go, which involves people searching for fantasy characters outside on the streets using their smartphone. There is the narrative of searching for Pokémon creatures and there is the narrative of the gamer crossing the street. Both narratives combine if and when the gamer crosses the street in order to look for the creature. Thus, if we consider the use, experience and phenomenology of the gamer (rather than taking a third person perspective), there is a sense in which there is *one* narrative.

In addition, this one narrative is connected with the narrative of the code running the application, which makes possible the game narrative. Indeed, it must be emphasized here that the narratives and the times of the gamers are configured by the technology; these ICTs are what Coeckelbergh and Reijers, influenced by Ricoeur, have called ‘narrative technologies’ (Coeckelbergh and Reijers 2016): like a text, they configure characters and events into a meaningful whole. The text of the code thus acts as a kind of author or at least co-author of the narrative of the gamer. But neither the narrative of the code nor the narrative of the gamer is more “real”, and the narrative of the game and the narrative of the gamer mix, without that it can be said that the one is more “real” than the other. One can use these terms from a third person perspective, of course, but if one tries to describe what happens by using the terms “real” and “illusion”, it is difficult to make sense of the experience of the gamer, who does not see his crossing of the street or his interaction with the robot as illusory; rather, there is one unified, integral experience. In the phenomenology of the game play (and the phenomenology of human-robot interaction, use of speaking computer interfaces etc.) there is no Platonic dualism; there is one game experience and one use experience.

Third, to further elaborate this approach one could also use the term *performance*. The metaphor of stage magic and illusionism of course already hints at performance. Indeed, Tognazzini writes:

Magicians work to produce illusions, but they don’t call their stage presentation an illusion, they call it an act (Tognazzini 1993, p. 356)

Now we can use this part of the metaphor and have it do some philosophical work, which again differs from the Platonic scheme Tognazzini uses. Following his own advice in the quote above, we can replace the language of reality and illusion by that of performance and act. There is one act, one performance. Or perhaps there are two performances, one done by the magician/designer and one done by the spectator and user, *who also performs*. Or, perhaps still better: there is one *co-performance*, in which both the designer/performer and the user/performer participate. Indeed, what is missing in the account presented by Tognazzini is the user/spectator in a more active role. In the Platonic cave metaphor/narrative, the prisoners are passive. They watch. They are even immobilized. Similarly, in Tognazzini’s account, the spectator is also relatively passive. In the magician’s show, spectators are literally immobile, sitting on a chair. With a few exceptions they do not participate in the performance. And it is assumed that the creation of the illusion is entirely taking place on the side of the magician. But this is misleading for at least two related reasons. First, consider again stage magic and illusionism, the metaphor itself. Performance can be seen as a one-way affair, but we can also take a different view, according to which the spectator does not passively receive meaning from the magician, but actively co-constructs time, narrative, and performance. In that sense, the spectator is indeed co-performer. Without the spectator, there is no act, no performance. To the extent that this is hidden by the metaphor of stage magic, the metaphor has its limitations. But an appropriate understanding of what is happening here and a better understanding of the metaphor itself reveals this more active role of the spectator. Second, consider now use of ICTs. The metaphor of stage magic almost hides that ICTs are *used* and that it is part of practices. Users are not (mere) spectators; they *do* something, they perform. The alternative approach I articulate here and its temporal, narrative, and performative turn away from Platonism, is only possible by considering ICTs in their *use* and *experience*. It is only by considering ICTS in their use and experience that the real/illusion is overcome. If we only look at *design*, as Tognazzini and others do, then we miss this aspect and easily remain within the Platonic framework, which in practice is shared by many designers, engineers, and scientists working in fields such as social robotics, engineering, and so on, but also by many philosophers. Then we see what is happening from a third person point of view. But we need to move beyond the language of “a view”, and especially an outside view. There are processes, there are narratives,

and there are performances, in which not only designers but also users are actively involved. Once we consider the performance and experience of the user of devices that “deceive”, the Platonic way of thinking evaporates. Then we see that there are techniques and technological artefacts used by designers, but also that there are techniques and artefacts used by the users of ICTs.

For example, if a person uses a robot as a companion, then we may distinguish between at least the following performances: there is the performance of the designer, who writes the code using a computer and computer programs, but there is also the performance of the user, who uses the robot, and there is the performance of the robot, which may use all kinds of artefacts. All these uses, experiences, and performances are part of a whole, they are connected through time and narrative, and through artefacts (especially the robot). There is also the performance of the company who wants data from the user. Now all these performances are “real”, and they involve various kinds of techniques, bodies, and artefacts. To describe what is going on *only* in terms of a deception designed by the designer/magician reduces a rich holistic performative configuration and process to only one performance, and—by using the term deception—gives ontological priority to one particular performance as opposed to others. Similarly, to focus on what the robot “really” “is” as opposed to the “illusion”, is to blind the analysis to all kinds of relations between this robot and various performances. The robot is embedded in narratives-in-the-making, there are processes and performances. If there is a reality at all, it is not a static “world” which we can “view” but a process-reality that is made and performed.

Moreover, as the “narrative technologies” approach mentioned already suggested, technology plays a more “active” role in these processes, narratives, and performances. Consider also Pickering’s reading of Latour (1993): Pickering argued that there is human and material agency (p. 21), that humans and machines ‘collaborate in performances’ (p. 16) and that there is ‘interplay here between the emergence of material agency and the construction of human goals’ (p. 56). This gives us a different view of stage magic and of the use of ICTs, in which the magician/designer is no longer totally in control of the performances. Instead, both users and machines co-write the narratives, co-configure the time/experience of the user, and co-perform. At a meta-level, then, instead of talking to the all-knowing Platonic philosopher in order to understand what is going on, we have to take advice from the users and performers: the magician/designer as user of technology and as performer, but also the users of technology, the performers of technology.

To conclude, according to this more holistic and relational alternative approach that takes a narrative and performative turn, instead of asking “What is real and what is illusory?” (the Platonic question), now the main question is

“What is going on?”, understood here as: what is going on in terms of time, narrative, and performance. This gives us a novel way of looking at the “deception” issue. If there is “deception” and “illusion” at all, it is a deception and illusion that are made in performances, and that are co-created and co-performed by humans (magician/designer and spectator/user) and non-humans (robots and other machines, artefacts, and devices).

This gives us a different approach to “deception” phenomena than, say Turkle’s or Sparrow and Sparrow’s, and suggests not only that we can dispense with a derogatory view of performance, but also that we do not need to use the language of deception. First, in contrast to Turkle’s use of the term, here performance is not seen as a negative term indicating illusion, but, decoupled from deception, magic, and illusion, it becomes a morally neutral process which involves humans and non-humans. Second, there may still be an ethical problem with robots that “pretend” having feelings, but this phenomenon and indeed problem, should not be framed in terms of pretence or illusion or “performance” as a negative and derogatory term, but in terms of performance as a morally neutral term to metaphysically bring together humans, and humans and non-humans. Humans and technologies co-perform and co-stage something here. Now in some cases this performance can rightly be seen as problematic: not because there are two different realities, but because there may be a problem with the performance and its consequences, or one could say that there are two conflicting performances. Let me unpack this.

What happens in so-called “deception” cases is that, on the one hand, the performance is successful, for instance in creating a robot with emotions. If the performance is successful, then in the experience of the viewer/user, there is not the “appearance” of emotions, there are emotions. On the other hand, at a different point in time or when viewed from the outside, the performance fails: it fails if and when others (e.g. philosophers or the same users at another time) think and say these emotions are not real (which is also a performance, one which uses language). Success and failure might also happen with regard to different groups at the same time. One group of users may experience the performance as successful, whereas another group of users may experience the same performance as unsuccessful. In such cases, instead of a deception problem, we have a performance problem: it is not entirely successful. This need not be problematic if everyone knows it was a show anyway; but it is problematic if the claim was that for instance the machine has emotions. Moreover, use of the language of deception is itself a performance, a counter-performance so to speak, which does not stand outside the performative field. In the case of the different user groups, one could say that there are different kinds of performances, and the term “deception” is used in a third performance to mark the difference between



the successful and the unsuccessful performance. One can also reframe the problem in narrative terms: what is missing here is not “the real” as opposed to “illusion”, but success or failure on the part of the designer and the robot—but also the user/spectator—to co-write a particular narrative, for instance a love narrative or a companion narrative. Or again there may be two conflicting narratives: one about love and one about deception. The ethical question then concerns the ethical quality and consequences of these performances and narratives (and indeed of this “battle” of performances and narratives) for the people involved. Is it good that young children get involved in a narrative of companionship with a robot? Is it good that a particular adult co-performs sex with a robot? Is it good that elderly persons with limited cognitive abilities become involved in a performance of care in which robots play a specific role? To answer these questions does not require a discussion of metaphysics or does not need framing in terms of deception; it requires us to attend to the specific human–robot interaction as a performative and narrative process in which the experience and co-performance of users counts. There may well be a difference between performances and also a difference in ethical quality, but that difference is a matter of (relational) phenomenology, of experience-in-relation and experience-in-performance; it is not a metaphysical or theoretical-scientific difference between what the robot can do and not do (the properties of the robot), the reality of the world, the nature of emotions, and so on.

### A relational turn

This move also invite us to connect with a relational approach to human-robot interaction (e.g. Coeckelbergh 2012, 2014), which enables us to criticize the distancing of the deception language. Those who use deception language or assume and real/illusion distinction tend to take what Nagel called a “view from nowhere”. While in general a third person point of view may not necessarily be problematic and probably is unavoidable, the *very distant* and detached view of the scientist qua scientist and philosopher qua metaphysician *is* problematic since it neglects the concrete relation between human and robot, human and human, and so on. By focusing on the properties of the robot (what the robot can do or cannot do, what the robot is or is not, has or has not, e.g. emotions or not), what remains out of sight is the concrete relation, encounter, (co-)performance, and experience. The ethical quality of the performance and whether or not it is successful is not a matter of what the robot is, has, or can do, or what the user/viewer/audience is, does, and so on, but of what happens in the relation between the two, here cast as: what happens during a (co-)performance. In the performative process and experience, there is no robot-in-itself and no human-in-itself; both

are co-constituted in the performance and in the relation. If there is a so-called “deception problem” then this must be understood as a relational problem: one which does not concern the robot but the relation between human and robot as performed. What happens in this relation needs ethical analysis and judgment. Moreover, such performance relations invite other performative interventions such as the voice of the designer-roboticist, the philosopher-ethicist, and so on who may or may not use the language of deception as a performative gesture—interventions which does not stand outside the performative field, and could themselves be criticized, for instance as involving too much distance. The so-called “deception” issue is then not about “the real” or about what emotions “are” but is rather a problem concerning the performances and narratives humans and robots create and should (not) create in specific cases, situations, and contexts, and about the ethical quality and consequences of these performances and narratives.

Thus, the advantage of this re-description and re-evaluation in performative and relational terms is that it is now possible to ethically evaluate the relational process, performance, and experience itself, indeed the relation itself and its consequences, without having to involve a third distant metaphysical entity such as “reality” (the real world, real emotions, etc.), the “nature” of emotions, etc. or abstract scientific-theoretical concepts such as anthropomorphism, which blind us to the quality concrete relation, encounter, and performance. Of course roboticists, human–robot interaction scholars, ethicists of robotics, etc. often also start from the concrete experience. But they then take distance and turn these into “cases” with their theories and generalizations. And when, *in their evaluative moments*, they use deception language or make Platonic assumptions, their use of language takes distance from the concrete performative-narrative and relational process, and to the extent that they do this, *their* performance becomes *itself* problematic.

These qualifications are important: the aim of my proposal to use a different language—that of performance, narrative, and relations—is *not* to discredit the work of authors such as Turkle, the Sparrows, and so on as invalid, entirely wrong-headed, and so on. Generally, these authors pay a lot of attention to the concrete human–robot interactions, especially Turkle for instance. My “only” suggestion is that those interested in better understanding and evaluating contemporary (social) robotics need to be careful and critical when using terms such as deception, when using scientific methods and theories, and when assuming metaphysical distinctions concerning the real etc., and consider using alternative terms that do more justice to performative and relational experience—experience on the part of all people involved. This includes users, but also designers/engineers, roboticists, philosophers, social scientists, etc., since they may *also* be involved and co-create different performances and narratives

with robots. Hence they cannot assume the role of “neutral” and distant observer; they themselves, with their science, criticism, interventions, gestures, etc. (which are also performances) influence and even co-constitute the performance, relation, and their meanings and ethical consequences.

To conclude, I have proposed to re-frame and further analyze the “deception” problem by using the notions of performances, narratives, and relations, which may help to avoid assumptions and discussions concerning for instance a metaphysics of the real, what robots can do, what emotions really are, etc. and which brings the discussion back to human-robot use and interaction as an experiential and performative-narrative process which must be analysed and evaluated in terms of the quality and consequences of its performances and experiences, including ethical quality. As scientist or philosophers we can do this from a third person point of view, if we must, but then we need to make sure that it is one that stays close to the phenomena and starts from there to develop a better understanding and, if necessary, an ethical judgment. And although an increasing amount of researchers may seem to do this, there is always the danger of a too distant theoretical or metaphysical attitude. Moreover, the arguments and discussion presented so far, with all its difficulties and potential pitfalls, suggests that we are only in the beginning of achieving a better understanding of the phenomena. I have argued that the moral language of “deception” and “illusion” may hinder rather than help in this process.

Now what does this approach mean for ethics of ICTs? And what does it imply for ICT design and use?

### **Implications for the ethics of, and responsibility for, ICT design and use: from an ethics of deception to an ethics of co-performance**

Let me first outline an ethics in the spirit of Tognazzini, which retains the language of deception, but uses deception in a less derogatory or negative way. Such an ethics would be a kind of virtue ethics, in particular an ethics of honesty. The ethics of honesty demands from designers to be honest about their robot, their speaking app, their game, etc. It requires from them, instead of hiding what they do, to design their robots in such a way that it is clear that it is a machine, a piece of technology, a game created by apparatus, etc. It requires them to make sure that there is a magician at work, who tricks users. The virtuous designer, then, is a Platonist magician who uses techniques and artefacts to entertain, but who knows the full truth and when the performance is over goes back into the cave to liberate the prisoners-spectators. Or rather, before imprisoning them, before the show starts, and perhaps even during the show, she makes sure that people are aware that all this is only trickery. This actually

requires from designers to take on a double role: that of the illusionist and that of the de-illusionist. On the one hand, the designer (and especially the company) needs to sell the device as magic. This is the attraction. It is also what people want. They want the show. For instance, someone using a virtual reality device to play games wants to be immersed in that game and want to be “fooled”. On the other hand, the designer also needs to reveal the tricks, or at least needs to reveal *that it is a trick*. Arguably this is often lacking in current robotics. Perhaps in the lab researchers are happy to show their tricks; but the robots sold on the market intentionally hide their tricks. They are meant to be magic, they are advertised to be magic, and generally the tricks are not visible and not transparent. Based on the ethics of honesty, one could demand more transparency.

This ethics is very attractive, and from this point of view it would be good if designers were to take it up, as compared to what, according to this approach, are the “lies” of social robotics, speaking machines, alternate reality games, etc. This ethics is compatible with an ethics of the right, which tells us that is not right to deceive. It is also compatible with a virtue ethics focused on honesty.

However, in the previous section I have complicated the picture. I have made it plausible that there is another, non-Platonic and less dualist and more relational way of thinking about the issue of deception and ICT, which calls for evaluating the use of (social) robots not in terms of “deception” but in terms of the success and ethical quality of the performances and their consequences. We should have a discussion about which narrative and performance we want (rather than about what is real and what is illusion) in which contexts and for which people. For instance, we may want robots to perform “machine” roles but not “lover” roles. We may want a performance of mowing the lawn, but not a performance of “friendship”. Or we may be fine with friendship with robots narratives for adults but not for young children. Once we drop the language of deception and the related real/unreal distinction, the ethical analysis must then evaluate not whether a particular phenomenon is real or not, but instead whether the performance, understood as a relational process, is good—in its success, quality, and consequences. As it stands, there seem to be good arguments why machines cannot be good performers when it comes to feeling, love, and all that is needed for good relationship performances and relationship narratives. It seems that they fail. Now at this point one may take again a more theoretical attitude and for instance consider Dreyfus’s philosophical criticisms of AI here or seek advice from the scientists: even within robotics and within a computational paradigm most roboticists recognize limits. For instance, Matthias Scheutz has argued that robots lack ‘the architectural and computational mechanisms that would allow them to care, largely because we do not even know what it takes, computationally, for a system to

care about anything’ (Scheutz 2011). However, these distant theoretical approaches are problematic; the “proof” here is not a matter of concepts or theory, but of performance. The proposed approach would frame these reasons not in terms of the *reality* or properties of the robot (its having emotions or not for instance) or in terms of what robots can and cannot do, theoretically speaking [to use Dreyfus’s famous title (Dreyfus 1992)]. It would re-frame these reasons in terms of its performance itself. Which “capacities” the robot has depends on performance, which is a relational and experiential matter: it always involves human subjectivity and experience, in relation to the artefact. What robots “are” depends on those relations and experiences, and the game of defining the properties and nature of machines is itself a particular performance, which should not escape criticism.

Thus, an ethics of performance is called for, which evaluates which performances and consequences of these performances are right and good—which then of course depends on how one defines right and good, or rather, how one experiences right and good. Designers, then, are responsible for the performances of the machines they design. And scientists and philosophers are responsible for *their* performances, which are not neutral, so-called “objective” interventions. However, when it comes to design one should not only put the responsibility on the part of the designers: when users and perhaps even machines are *co-performers*, then ethics can no longer be one that assumes that full responsibility lies only with the magician/designer. Indeed, if we follow this alternative route I think there are at least the following implications for the *responsibility* with regard to ethics of ICTs, which now can be understood as concerning an ethics of deception rather than an ethics *against* deception, or better: an ethics of performance rather than an ethics against performance. Let me unpack these claims.

First, designers of ICT are responsible for their design performances, which result in uses of artefacts and performances (by the artefacts they designed and by users) and in specific narratives and configurations of the time of the users. The ethical question then is no longer “Are you honest about what you do?” but is rather: “What kind of narratives do you create and what kind of performances do you enable by designing the technology in such and such a way?” And what are the short-term and long-term consequences of these narratives and performances? These narratives and performances can then be evaluated by ethical criteria and theories, for instance by means of a virtue ethics, where the question is: “What kind of characters—human and non-human—and what kind of events and plots do you create?” Designers of the ICTs considered in this paper should then be held responsible, not so much for how honest they are with regard to what they are doing but rather for the ethical consequences of their designs for the characters and lives of people, which as uses of technology, as experiences, as

(co-)performances, and as narratives, are configured and reconfigured by the technologies they designed. The main ethical problem and challenge directed at the designer/magician is then not “Is this a show or not, is this trickery or not? Are you fooling us?!”; we generally know and accept that there is a show, we know that there is trickery, we know that human social life is a performance and that use of technology is a performance. Rather, the question with regard to ICT design is: “Given that you use all these tricks and techniques, given that you co-shape our performances and our narratives, given that the shows goes on and that the show *must* go on (given that, in this view, there is no “outside” to performance and no outside to technology-in-use), how can we ensure that these narratives and performances and their consequences are good—according to a particular definition of “good”?”

Indeed, let me note again that the approach presented here is compatible with an ethics of good and an ethics of the good life, and is open to various definition of “good” and the “good life”, provided that theories involved are compatible with a relational, narrative and performative turn. It is assumed here that “good” is not something that is detached from experience and wisdom in concrete practical performances and lived narratives. It is also compatible with narrative ethics and virtue ethics, but then a virtue ethics that does not only exclaim that we should be honest, but re-defines honesty and other virtues starting from the assumption that design is about performance, and that character is “made” in such performances and not merely the result of human will or intention; it is also co-configured by the technologies and techniques used in these performances (more work is needed to develop this point).

Second, however, since as argued users co-write these narratives and co-perform them, since they are part of the magic performance as co-authors and indeed co-magicians, they are *also* responsible. Design is often defined in terms of intention, plan, aim, and so on (Flusser 1999, p. 17). But if the magician/designer is no longer in full control of the show and if the show crucially depends on the users as co-performers, the responsibility of the magician/designer is limited, both in terms of the success of the performance and in terms of the ethical quality of the performance. If we no longer see users as passive receptors of whatever the designer has in mind and intends, then they should also be held responsible for the show. Moreover, that responsibility is *also* limited: *both* the designer and the user have to accept that the resulting performances and narratives are not entirely under their control. As I indicated before, there may be unintended consequences. For design, this means that the design should be “open” enough to enable and allow different kinds of uses (that is, co-performances by users), and that it should take into account unintended consequences as much as possible, for

instance by means of creating and evaluating all kinds of (worst case) scenarios.

Moreover, a relational view also means that we should look beyond individual users as co-performers: we should also consider relations within families, groups, wider society. Performances are always part of a larger whole, and are normatively embedded in, and shaped by, these larger wholes (see also Coeckelbergh 2017). Here too there could be unintended consequences for, say, society as a whole—consequences which are normatively relevant. In the last section, I will say more about the politics of ICT design.

Finally, these unintended consequences have at least partly to do with the more “active” role of technology. If, in line with contemporary philosophy of technology including the “narrative technologies” approach mentioned here, in line with approaches in STS such as those inspired by Latour’s work, and *in contrast* to Tognazinni and his many colleagues—designers and entrepreneurs, see technology not as “passive” artefacts that are mere instruments of the will and intention of the magician/designer but rather as co-performers and co-authors of the magical narratives and illusionist performances under considerations, then this raises the question if machines can also be co-responsible. Through the magic, machines increasingly appear as such. In practice, people will be increasingly drawn into performances that tempt them to acknowledge the virtual responsibility (Coeckelbergh 2009) of machines. However, it would be ethically problematic to confuse the performance of agency with the performance of responsibility; the latter, if attempted, fails in the case of robots. When machines try to do a “responsibility” performance, they fail. We experience that failure. We do not receive a response. The machine performs a reaction.

That being said, the role of the mentioned technologies is more than that of passive instruments; they also “act” in all kinds of ways—with Akrich and Latour we could call them ‘actants’ (Akrich and Latour 1992, p. 259)—even if they cannot pull off the act of response (as opposed to reaction). Humans must remain responsible. In contrast to Latour, who has a symmetrical view, I propose an asymmetrical view: humans are the main performers, even if they are not the only performers. But as co-performers and as experiencers of the performance—as designer or as user—they are crucial and irreplaceable. They are always part of the relation. To fully support and justify this view, however, would lead me into a different discussion; it would require another philosophical performance.

For now I conclude that the ethical discussion of cases and ICTs of “deception” could be enhanced by looking at it through the lens of stage magic and illusion, that philosophers of technology could learn from relevant work on this by designers and scientist in the field of ICT, that this angle leads us to consider large metaphysical and ethical

questions, and that this paper has mapped and articulated two positions, which involve different approaches to the question concerning the real and which enable us to ask different and interesting ethical questions about the magic and trickery of contemporary and near-future ICTs such as companion robots, talking machines, and games and other technologies that leave the “virtual” world and enter the streets.

Let me now end this paper by trying to further think through the approach explored and articulated in this paper, and explore what it means for politics, in particular the politics of ICT design and use.

### **Illusio ex machina and the political promise of Ulysses’ trickery: political deception and political performances**

Does the previous discussion ask for a ban on using the term deception? Maybe the term can still be “saved” if we release it from its metaphysical and theoretical burdens, and interpret it in a performative way according to the approach I proposed. Then we can say that we need an ethics of deception, interpreted not as one that is concerned with reality versus illusion, but one that is about the right and good performance. Then we can keep the alternative approach, but still use the term deception if we want, for instance to make the link with magicians’ practices. However, to avoid confusion with the dualist and metaphysical approach I criticized, I propose to use the term performance.

This issue is also relevant when we consider the politics of design, to which I now turn. One reason for why it might be wise not to entirely ban the term “deception” from a discussion of design of ICTs, in particular design of *machines*, is that the meaning of “machines” is etymologically related to deception. In his philosophy of design, Flusser (1999) reminds us that the noun design is often connected with cunning and deception—a connotation which interestingly also applies to the term *machines*:

The word occurs in contexts associated with cunning and deceit. A designer is a cunning plotter laying his traps. Falling into the same category are other very significant words: in particular, mechanics and machine. The Greek *mechos* means a device designed to deceive—i.e. a trap—and the Trojan Horse is one example of this. Ulysses is called *polymechanikos*, which schoolchildren translate as ‘the crafty one’. The word *mechos* itself derives from the ancient MAGH, which we recognize in the German *Macht* and *mögen*, the English ‘might’ and ‘may’. Consequently, a machine is a device designed to deceive; a lever, for instance, cheats gravity, and ‘mechanics’ is the trick of fooling heavy bodies (Flusser 1999, p. 17)

The design of machine technology, then, can itself be defined as the art (*techne*) of deception. The ethical question is then not *if* designers should deceive, it seems, but rather what kind of traps, cheats, deceits, and fooling we need and want—to deceive ourselves and have ourselves deceived. This is in line with the alternative approach I explored in this paper, at least if we then take the step of re-casting this deception narrative in terms of *performance*: a designer is a craftsperson who designs machines—this design is by itself a performance—that enable a particular kinds of performances with nature and with others. A good designer, then, is a cunning performer who enables others to do tricky performances.

Moreover, Flusser’s etymological note also raises the question of power: the magician/designer is also the one who is able to do something, perform something, who has the power to do something—with power understood as a capacity and a potential (in German: *Vermögen*), in particular here: the capacity to perform. This raises the *political* questions regarding the design of technology. Put in the language of deception, we can ask: given that design is all about deception, who has the capacity to deceive, who has *more* capacity to deceive than others, and who deceives whom? Who should have the capacity and opportunity to deceive? Put in the language of performance, this becomes: who performs with whom, and who lets other perform what? Who has how much capacity and power, the *Vermögen*, to perform, in particular contexts and with regard to particular aims and projects?

Now when it comes to a normative position, the language of deception once again offers the temptation to revert back to the ethics of honesty, and avoid these questions by saying that one should not deceive in the first place. The ethics of honesty seems to require a politics of transparency and de-masking, which is aimed at revealing the tricks and deception of the designers and politicians, and at democracy as requiring, among other things, to make all knowledge involved transparent. Again, there is a lot to say for such an approach and such a normative orientation. It is in line with the intuitions of many—including some of my own. However, if we start from the different assumptions and approach articulated in this paper, what are the consequences for thinking about the politics of ICTs and, more generally, the politics of technology?

If we acknowledge, with Flusser, that technology deceives, and if we acknowledge that technology is unavoidably used for manipulation of other people, then we may ask different questions. If technology is defined as a psychological weapon to exert control over other human beings (Gell 1994, p. 1988), or if we accept, perhaps inspired by Foucault (1975), that there is always going to be disciplining in society, then we may still ask ethical and political questions: questions which do not concern *if* there is deception and

manipulation or if deception in general is bad, but rather ask who deceives and manipulates whom, which deceptions are good, and put in the language of performance: who manages to have others perform for them, and which kind of manipulative and disciplining performances are acceptable and good.

There are of course plenty of criteria in moral and political philosophy for what is acceptable and good, and some of them have been applied in robot ethics. For instance, Sparrow (2016) has evaluated technology in aged care according to what he believes are objective elements of welfare. All these criteria and discussions are not rendered obsolete by the approach I propose. But now they perform a new function: as criteria to evaluate performance. For instance, one may ask: Is the co-performance of a particular health care robot contributing to the welfare of the residents of this elderly care home, according to criteria for welfare as defined by Sparrow? Then the question is not about deception but about the quality and ethical consequences of the performance. These criteria may allow for some kinds of manipulations but not for others.

To show what this means in the context of political principles, let me here focus on the political ideal of democracy. The alternative ethics presented in this paper, when connected with a democratic ideal, suggests a different normative position, which does not avoid but answers the question “who should have the capacity and opportunity to deceive?” The democratic answer here, it seems, is that not only responsibility but also *power* should be distributed to both designers and users. If we are going to have to play the game of deception at all, so this position goes, then users should also have a hand in it. Of course, if users already play the game and co-perform, as I argued, then in a sense they already have power, then all of us already have power. If they had no power at all, they could not exercise their responsibility, and indeed the question of responsibility would not come up in the first place. Yet while this is the case for all users, it is more the case for some users than for others. For instance, people who know how to code seem to have more design power (understood as deception power or manipulative power) and more performance power with regard to ICTs than others, based on their skills. And today there are many users/designers, “hackers”, who assume more power than initially given to them by the designers—again since they are more skilled, have more know-how to deceive than others. They know how to design their Trojan horses. Thus, there is still inequality in this respect; some people have more skills and hence power than others. Now one could say that it is up to users to acquire more skills. But neither technology nor its social context should be seen as a given. There may be power structures in society that give more opportunities to some people than to others. Some have more power to do some performances or to have others perform for them.

And even if we all had the same skills, there might be still social mechanisms through which some people get more opportunities and more power to deceive/design than others. Perhaps it should be a meta-ethical duty for designers and politicians to design technologies and social structures that create enough room for the self-empowerment of *all* persons as citizens-users. The guiding normative-political ideal here, again based on the ideal of democracy and the alternative approach articulated in this paper, is that deceit is only problematic if it is in the hands of one person or one group of people. Deceit and self-deceit by means of technology and by other means, then, is not a problem as long as the power to do so is democratized (and as long as ethical criteria are satisfied). The political promise of the democratic ideal is then translated into the hope that we can all become better, more skilful and more powerful magician-designers and magician-performers, that we can all become Ulysses. Alternatively, if we drop the vocabulary of illusion and deception and leave an individualist framework, we could say that the political promise is that we can all become better (co-)performers, and that we can better perform together. Whether or not we call it deceit or trickery, there is not only the option of the lone hero; there is also the possibility of achieving the cunning intelligence of collaborative performance.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Adar, E., Tan, D. S., & Teevan, J. (2013). Benevolent deception in human computer interaction. In *CHI 2013*, ACM (conference paper), April 27–May 2, 2013, Paris, France. Accessed February 4, 2017, from <http://www.cond.org/deception.pdf>.
- Akrich, M., & Latour, B. (1992). A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies. In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 259–264). Massachusetts: MIT Press.
- Ananthaswamy, A. (2016). *Virtual reality could be an ethical minefield: Are we ready?*. New Scientist, 4 March 2016. Accessed February 4, 2017, from <https://www.newscientist.com/article/2079601-virtual-reality-could-be-an-ethical-minefield-are-we-ready/>.
- Coeckelbergh, M. (2009). Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents. *AI and Society* 24(2):181–189.
- Coeckelbergh, M. (2017). *New romantic cyborgs: Romanticism, information technology, and the end of the machine*. Cambridge, MA/London: The MIT Press.
- Bryson, J. (2010). Robots should be slaves. In Y. Wilks (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues* (pp. 63–74). Amsterdam: John Benjamins.
- Carpenter, J. (2016). *Culture and human–robot interaction in militarized spaces: A war story*. New York: Ashgate.
- Coeckelbergh, M. (2012). *Growing moral relations: Critique of moral status ascription*. New York: Palgrave Macmillan.
- Coeckelbergh, M. (2014). The moral standing of machines: Towards a relational and non-cartesian moral hermeneutics. *Philosophy & Technology*, 27(1), 61–77.
- Coeckelbergh, M. (2017). Technology games: Using Wittgenstein for understanding and evaluating technology. *Science and Engineering Ethics* pp. 1–17.
- Coeckelbergh, M., & Gunkel, D. (2014). Facing animals: A relational, other-oriented approach to moral standing. *Journal of Agricultural and Environmental Ethics*, 27(5), 715–733.
- Coeckelbergh, M., & Reijers, W. (2016). Narrative technologies: A philosophical investigation of narrative capacities of technologies by using Ricoeur's narrative theory. *Human Studies*, 39, 325–346.
- Dreyfus, H. (1992). *What computers still can't do*. New York: MIT Press.
- Floridi, L. (Ed.) (2015). *The onlife manifesto*. Cham: Springer
- Flusser, V. (1999). *Shape of things: A philosophy of design*. London: Reaction Books.
- Foucault, M. (1975). *Discipline and punish: The birth of the prison*, Trans. A. Sheridan. New York: Vintage Books/Random House.
- Gell, Al. (1994). The technology of enchantment and the enchantment of technology. In J. Coote (Ed.), *Anthropology, Art, and Aesthetics*. Oxford: Clarendon Press.
- Kidd, C. D., Taggart, W., & Turkle, S. (2006). A sociable robot to encourage social interaction among the elderly. In *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, Orlando, Florida, May 2006, pp. 3972–3976.
- Latour, B. (1993). *We have never been modern*. (trans: Porter, C.) Cambridge, Massachusetts: Harvard University Press.
- Marshall, J., Benford, S., & Pridmore, T. (2010). Deception and magic in collaborative interaction. In *CHI 2010: Performance, stagecraft, and magic*, ACM (conference paper), Atlanta, GA, USA, April 10–15. Accessed February 4, 2017, from <http://www.cs.nott.ac.uk/pszjm/2/wp-content/uploads/2010/04/marshall-chi-deception.pdf>.
- Musial, M. (2016). Magical thinking and empathy towards robots. In J. Seibt, M. Nørskov, & S. A. Søren (Eds.), *What social robots can and should do. Proceedings of Robophilosophy 2016*. Amsterdam: IOS Press, pp. 347–355.
- Pasquinelli, E. (2010). The illusion of reality: Cognitive aspects and ethical drawbacks: The case of second life. In C. Wankel, & S. Malleck, (Eds.), *Emerging ethical Issues of life in virtual worlds*. Charlotte, North Carolina: IAP
- Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago: University of Chicago Press.
- Rowe, N. C. (2007). Deception in defense of computer systems from cyber-attack. In L. J. Janczewski & A. M. Colarik (Eds.), *Cyber war and cyber terrorism* (pp. 97–104). New York: Information Science Reference.
- Scheutz, M. (2011). The inherent dangers of unidirectional emotional bonds between humans and social robots. In P. Lin, G. Bekey & K. Abney (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 205–222). Cambridge, MA; London: MIT Press.
- Sharkey, N., & Sharkey, A. (2010). The crying shape of robot nannies: An ethical appraisal. *Interaction Studies*, 11(2), 161–190.
- Sparrow, R. (2016). Robots in aged care: A dystopian future? *AI & Society*, 31, 445–454.
- Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, 16(2), 141–161.
- Sparrow, R., Harrison, R., Oakley, J., & Keogh, B. (2015). *Playing for fun, training for war: Can popular claims about recreational video gaming and military simulations be reconciled?* *Games and Culture*. Published Online First, November 26, 2015.
- Sung, J.-Y., Guo, L., Grinter, R. E., & Henrik, I. C. (2007). “My Roomba is Rambo”: Intimate home appliances. In J. Krumm et al. (Eds.) *Proceedings of UbiComp 2007, LNCS 4717, Lecture notes in computer science 4717* (pp. 145–162). Berlin: Springer.
- Tognazzini, B. (1993). Principles, techniques, and ethics of stage magic and their application to human interface design. In *INTERCHI'93*,

*ACM (conference paper)*, pp. 355–362. Accessed February 4, 2017, from <http://dl.acm.org/citation.cfm?id=169284>.

Turkle, S. (2010). In good company: On the threshold of robotic companions. In Yorick Wilks (Ed.), *Close engagements with artificial companions*. Amsterdam: Basic Books.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.